البحث الثالث

المقارنة بين ثلاث طرائق في تقدير ثبات الاختبارات المركّبة التي تتضمن نوعين من الفقرات (ألفا ، ألفا الطبقي، راجو)

د. حسان غازي العمري *

الملخص

هدفت هذه الدراسة إلى المقارنة بين بين ثلاث طرائق في تقدير ثبات الاختبارات المركبة التي تتضمن نوعين من الفقرات باستخدام أسلوب كل من (ألفا، ألفا الطبقي، راجو) والتي يستخدم كل منها في تقدير ثبات الاختبارات المركبة المتشاكلة congeneric tests وذلك باستخدام بيانات 84 اختباراً من الاختبارات النهائية التي أعدها أعضاء هيئة التدريس في جامعة جدارا تم اختيارها عشوائياً من المساقات الدراسية المطروحة في مرحلة البكالوريوس. إستجاب على كل اختبار في المتوسّط حوالي 30 طالباً وطالبة وكان متوسّط عدد الفقرات حوالى 13 فقرة في كل اختبار وتكونت جميعها من نوعين فقط من الفقرات هما: الاختيار من متعدد والمقالية. أشارت نتائج الدراسة إلى أن الوسط الحسابي لتقديرات الثّبات قد بلغ (0.71) (0.79) (0.53) حسب معادلة ألفا وألفا الطبقي وراجو على الترتيب. كما أشارت نتائج اختبار تحليل التباين ANOVA واختبار كاي تربيع إلى وجود فروق دالة احصائيا بين توزيعات معاملات الثّبات وبمعامل توافق بينها Contingency coefficient بلغ 0.36.

الكلمات المفتاحية: الثّبات،الاختبارات الجامعية، الاختبارات المتشاكلة، راجو، ألفا، ألفا الطبقي.

* أستاذ مساعد في قسم القياس والتقويم – كلية العلوم التربوية – الجامعة الأردنية.

1- خلفية الدراسة:

يعتبر الثّبات من أهم الخصائص السيكومترية التي يجب أن يتمتع بها الاختبار، فالثّبات يعني الدقّة في القياس، ولكن تقدير هذا الثّبات هو بحد ذاته مشكلة سيكومترية معقدة، فقد تعددت محاولات وأساليب تقديره بدءاً من تقدير الثّبات بالإعادة، ثم استخدام النماذج المتوازية، إلى فكرة استخدام الاتساق الداخلي والتي تقدم تقديراً لثبات الاختبار من النتائج على نصفيه X1 و X2 وفق المعادلة 1:

$$\rho XX' = 2\rho X1X2 / (1+\rho X2X1) \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad 1$$

حيث: $\rho XX'$: معامل ثبات الاختبار بالتجزئة النصفية

$\rho X1X2$: معامل الارتباط بيرسون بين نصفي الاختبار X1 و X2

حيث تستخدم شريطة أن يتحقق في النصفين شرط التوازي والمتمثل بافتراض تساوي تباين العلامة الحقيقية وتساوي تباين الخطأ على جزأي الاختبار (Haertel, 2005).

وكون افتراض التوازي شرط من الصعب تحقيقه، فقد قُدّمت العديد من المعادلات الأخرى لحساب الثّبات باستخدام التجزئة النصفية بتوفر شرط أقل تشدداً من شرط التوازري سُمّي شرط التكافؤ بالضرورة (Essential τ - Equivalence) والمتمثل بـ (تساوي التباين الحقيقي لجزأي الاختبار بفرق ثابت دون تساوٍ في تباين الخطأ) (Lord & Novick, 1968) . ومن أشهر هذه المعادلات معادلة فلانجان (Flanagan, 39) التي تقدر الثّبات وفق المعادلة :2

$$\rho XX' = 4covx1x2 / \sigma 2X) \ldots\ldots\ldots\ldots\ldots \quad 2$$

حيث : covx1x2 التباين المشترك على جزأي الاختبار

$\sigma 2X$: تباين الدرجة الكلية على الاختبار

كذلك معادلة رولون (Rulon, 39) ومعادلة جتمان (Guttman, 45) المتكافئتان وفق الصيغة (3) أدناه:

$$\rho XX' = 2(1- (\sigma 2X1 + \sigma 2X2) / \sigma 2X2) = 1-(\sigma 2X1 - X2/\sigma 2X) \ldots 3$$

حيث: $\sigma 2X1$: تباين الدرجة الكلية على الجزء الأول من الاختبار

σ2X2: تباين الجزء الثاني من الاختبار

σ2 X1-X2: تباين الفروق على نصفي الاختبار

ومع أن شرط التكافؤ بالضرورة اقل تشدداً من شرط التوازي الاّ أن قضية تحقيقه بقيت أمراً صعباً إلى أن ظهرت فكرة جديدة تم فيها اعتبار أجزاء الاختبار أجزاءا متشاكلة (congeneric) وهي التي (تسمح باختلاف التباين الحقيقي والتباين الخطأ على جزأي الاختبار) وهو شرط اكثر معقولية، إذا أن تحقيقه أسهل بكثير من تحقيق شرطي التوازي والتكافؤ، وقد قدمت في هذا الاتجاه الكثير من المعادلات التي تفترض تشاكل الأجزاء، ففي حالة تجزئة الاختبار إلى جزأين، قدم راجو (Raju, 1977) اسلوباً لتقدير الثّبات عند معرفة اطوال الأجزاء، فإذا كان طول الجزء الأول λ1 فإن طول الجزأ الثاني (λ1 -λ2=1) بشرط أن 2(λ2 / λ1) = (σ2T1 / σ2T2) وعليه تكون معادلة تقدير الثّبات وفق الصيغة في المعادلة (4):

(ρXX ′= covx1x2 / λ2λ1σ2X) 4

مع الإشارة إلى أن قيمة هذه المعادلة مساوية لقيمة المعادلة التي قدمها فلانجان (Flanagan, 39) إذا كانت λ1 =λ2= 0.5 ، بعد ذلك قدم فيلت (Feldt & Brennan, 1989) أسلوباً في تقدير الثّبات معتمدا أسلوباً آخر في تقدير قيم λ1 و λ2 حيث تتم حساب قيمة كل منهما وفق المعادلة (5):

(λ2 =1- λ1) و λ1 = (σ2X1 + covx1x2 /σ2X)..... 5

أما في حالة التجزئة إلى اكثر من جزأين، فقد قدمت معادلة كرونباخ ألفا والتي تقوم على فكرة تجزئة الاختبار إلى N من الأجزاء المتوازية وتعطى بالمعادلة 6 وتساوي قيمته لمعامل هويت (Hoyt) الذي يتم تقديره من تحليل التباين (Crocker, 1987)، بالإضافة إلى إمكانية استخدام معادلة سبيرمان- براون العامة (6):

ρXX′ = n (ρXX′)/(1+(n-1) ρXX′).....6

أما إذا كانت الأجزاء ليست متوازية وليست متكافئة بالضرورة فإن قيمة كرونباخ ألفا تعتبر حداً أدنى للثبات (Novick & Lewis,67). بالإضافة لذلك وفي حالة الأجزاء المتشاكلة فقط، فقد قدم راجو Raju تعميما لمعادلته السابقة لتصبح صالحة للاستخدام فيما لو كان الاختبار مكوناً من ثلاثة اجزاء متشاكلة فأكثر معروفة الأوزان على الشكل الوارد في المعادلة (7):

$$R = \left(\frac{1}{1-\sum \lambda_i^2}\right)\left(1-\frac{\sum \sigma_i^2}{\sigma_x^2}\right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 7$$

وتستخدم إذا كانت أجزاء الاختبارات متشاكلة (congeneric)، وهو المفهوم الذي قدمه جورسكوج (Joreskog, 1971) حيث أشار إلى أن التشاكل يسقط فيه شرطي تساوي التباين الحقيقي وتساوي تباين الأخطاء، وبذلك تكون العلامات الحقيقية في النماذج المتشاكلة مرتبطة خطياً مع عدم ضرورة تساوي الأوساط والتباينات فيها، تلاها بعد ذلك المعادلة التي قدمها فيلت وبرنان Feldit & Brennan,1989)) في حالة التشاكل الكلاسيكي للأجزاء غير معروفة الأوزان.

2- مشكلة الدراسة:

يعتبر معامل الثّبات كرونباخ ألفا الأكثر والأوسع استخداماً في الأدب التربوي وعلم النفس(Kamata, Turhan, Darandari, 2003) بالرغم من أن استخدامه غير مناسب في كثير من الأحيان (Qualls,1995) وخصوصاً في حالة الاختبارات التحصيلية التي تتضمن في أكثر من نوع واحد من الفقرات (إذ تبرز مشكلة اختلاف الأوزان على هذه الفقرات الذي يجعل تحقيقها لشروط تقدير الثّبات باستخدام معادلات الاتساق الداخلي سواء في المعادلات التي تشترط التوازي أو التكافوء بالضرورة أمرا صعباً) الاّ أنها تستخدم بانتهاك افتراضها الأساسي المتمثل بتوازي الأجزاء ووجود ارتباطات تامة بين العلامات الكلية على هذه الأجزاء مما يؤدي استخدامها إلى الحصول على قيمة لكرونباخ ألفا تبخس معامل الثّبات (under estimate) وتمثل الحد الأدنى له وهو ما أشار له (Novick & Lewis, 1967).

وفي ضوء عدم إمكانية الحصول على أجزاء متوازية أو متكافئة، فإن افتراض التشاكل (Congeneric) للعلامة الكلية على الاختبارات التي يعدها المعلمون واعتبارها علامة مركبة من عدة

علامات فرعية أمر اقرب إلى الواقع بدلاً من افتراض توازيها أو تكافؤها، لأن شروط التشاكل (Congeneric) أقل تشدداً إذ تشترط فقط أن تكون الأجزاء مرتبطة خطياً مع امكانية عدم تساوي الأوساط والتباينات (كما تم الإشارة لذلك سابقا) وقد قدمت العديد من المعادلات لتقدير الثّبات بهذا الأسلوب (التشاكل) مثل معادلة راجو (Raju, 1977) ومعادلة فيلت (Fieldt, 1975) ومعادلة انجوف – فيلدت (Qualls ,1995) بالإضافة إلى معادلة ألفا الطبقي (Stratified Alpha) التي تتعامل مع الاختبار على أنه مكون من عدة طبقات تمثل فيه كل طبقة نوعاً من أنواع الفقرات(Brennan, 2006).

هذه المعادلات المتعددة لم تحظ بدراسات للمفاضلة بينها في تقدير قيم الثبات أو حتى المقارنة بين قيم التقديرات التي تقدمها وخصوصاً في ظل بيانات حقيقية من الميدان (Real Life Data)، ولذلك تجيء هذه الدراسة لتوفر معلومات حول الفروق في قيم التقديرات التي تقدمها ومحاولة تقصي أسبابها. وهي الدراسة الوحيدة التي تبحث في العلاقة بين معادلات تقدير الثّبات الثلاثة سالفة الذكر حيث اكتفت معظم الدراسات السابقة بتقديم طرائق لتقدير ثباتها مباشرة دون الكشف عن مدى التوافق بين هذه الطرائق.

3- أهمية الدراسة:

لهذه الدراسة أهمية من الناحيتين النظرية والتطبيقية، حيث تعتبر نظرياً الدراسة الأولى التي تقدم مقارنة بين هذه الطرائق الثلاثة من طرائق تقدير الثّبات وتبين مدى التوافق بين معاملات الثّبات المقدرة وفقها وذلك في ضوء بيانات حقيقية من الميدان (Real Life Dtata). أما من الناحية التطبيقية فهي تلفت نظر المتختصين في التربية إلى الاستخدام الخاطئ والشائع لمعادلة كرونباخ ألفا في تقدير ثبات الاختبارات التحصيلية وتقدم لهم بديلاً أفضل منها لاستخدامه في تقدير الثّبات في حالة الاختبارات التي تتكون من نوعين فأكثر من الفقرات. كما وتقدم معلومات حول مستويات الثّبات للاختبارات التي يعدها أعضاء الهيئة التدريسية في الجامعات بشكل عام، وفي جامعة جدارا بشكل خاص.

4- أسئلة الدراسة:

تتمثل مشكلة الدراسة في الإجابة عن السؤالين التاليين:

4- 1- ما تقديرات الثّبات للاختبارات الجامعية التي تتكون من نوعين من الفقرات بإستخدام كل من الطرائق الثلاثة: ألفا، راجو، وألفا الطبقي؟

4- 2- ما مدى التوافق بين كل من الطرائق الثلاثة (ألفا، راجو، وألفا الطبقي) في تقدير الثّبات لهذه الاختبارات التي تتكون من نوعين من الفقرات؟

5- مصطلحات الدراسة:

5- 1- طريقة كرونباخ ألفا: قدمها كرونباخ عام 1951 واصبحت الأكثر استخداماً حيث يرد توثيقها في الابحاث عدة مئات من المرات سنويا(Cronbach, 2004, Hogan, Benjamin & Brezinzki, 2000). تستخدم هذه المعادلة إذا كان الاختبار مكوناً من أكثر من ثلاثة أجزاء، وكانت الارتباطات بين العلامات تامة على الأجزاء المختلفة (Croker and Algina ,1998) وقد أجاز (Frisbie, 1988; Keng, Miller, O'Malley,Turhan,2005) استخدامها كطريقة لتقدير معامل الثبات للاختبارات التحصلية المكونه من نوعين من الفقرات وهي كما في المعادلة 8 أدناه.

$$\alpha = \frac{n}{n-1}\left[1 - \frac{\sum {\sigma_{xi}}^2}{{\sigma_x}^2}\right] \quad \ldots\ldots\ldots\ldots 8$$

حيث n عدد الفقرات في الاختبار، σ2Xi التباين الظاهري لكل جزء أو لكل فقرة، σ2X تباين الدرجة الكلية على الاختبار.

5- 2- طريقة ألفا الطبقي: وهي الطريقة التي قدمها كل من جلسر وكرونباخ وراجارتنام (Cronbach Gleser, & Rajaratnam, ,1965) وتعطي هذه الطريقة تقديراً للثبات يتعامل مع الاختبار كجزأين أو أكثر منفصلين، وفق المعادلة 9 التالية:

$$_{strata}\rho_{XX'} = 1 - \frac{\sum {\sigma^2}_{X_i}\left(1 - {}_a\,\rho_{X_iX_j}\right)}{{\sigma^2}_X} \quad \ldots\ldots\ldots 9$$

طريقة راجو: وهي الطريقة التي قدّمها قدم راجو Raju لتقدير الثّبات في فيما لو كان الاختبار مكوناً من ثلاثة أجزاء متشاكلة فأكثر معروفة الأوزان وقد تم توضيح هذه الطريقة كما ورد في المعادلة 7 أعلاه.

6– دراسات سابقة:

الدراسات في موضوع المقارنة بين أساليب تقدير معامل الثّبات قليلة، وخصوصاً المفاضلة بينها في تقدير ثبات الاختبارات التحصيلية، فقد أجرى كاماتا وتورهان وداراندارى (Kamata et al., 2003) دراسة هدفت إلى مقارنة اربعة أساليب في تقدير الثّبات في ضوء بيانات مولدة أظهرت ارتفاعاً في تقديرات ألفا الطبقي نسبة لغيره من الأساليب الأخرى، ودراسة كل من كورتينا وشيمت واوسبين (Cortina, 1993; Schmitt, 1996; Osbourn, 2000) والتي أظهرت إنحاساً في تقدير معامل الثّبات ألفا خصوصا إذا كان الاختبار متعدد الأبعاد، كما وأنه يُظهر تحيّزاً في تقدير الثّبات إذا كانت الفقرات تقيس ابعاداً مختلفة (Zimmerman et al., 1993; Murphy & DeShon, 2000) أمّا دراسة اوسبورن (Osbourn ,2000) فقد أظهرت أن استخدام كل من أسلوبي كرونباخ ألفا وألفا المعياري Standardized alpha يبخسان قيمة الثّبات المقدّرة في حالة تعدّد ابعاد المقياس حتى وإن كان عدم التجانس متوسّطاً. كما أظهرت الدراسة أن ألفا الطبقي يحسّن من معامل الثّبات في حالة تكوّن الاختبار من نوعين من الفقرات.

في الدراسات العربية لم تجر أي دراسة للمقارنة بين طرائق الثّبات إلا أن هناك دراستان استخدم فيهما معامل راجو (Rajo,1977) لتقدير ثبات العلامات الجامعية وليس الاختبارات الجامعية، فقد أجرى السوالمه (1995) دراسة هدفت إلى تقدير ثبات العلامات في الجامعية في عينة من المساقات في جامعة اليرموك باستخدام معادلة راجو وقد أشارت نتائج الدراسة إلى تمتع العلامات في 56% من المساقات الجامعية بدرجة مقبولة من الثّبات، حيث يزيد معامل الثّبات لها عن 0.70. وأجرى الشايب (2007) دراسة هدفت إلى تقدير ثبات العلامات النهائية في جامعة آل البيت، وقد كشفت النتائج باستخدام أسلوب راجو لتقدير الثّبات عن تدنّي معاملات ثبات العلامات بشكل عام، حيث بلغ المتوسّط العام لتقديرات الثّبات (0.66) كما أشارت النتائج إلى أن (31.25%) فقط من المواد تتمتّع

بمعاملات ثبات مقبولة كما أشارت نتائج التحليل الإحصائي مربّع كاي (χ^2) للاستقلالية إلى عدم وجود علاقة بين ثبات علامة المادة ومستوى المادة أو الكلية التي ينتمي لها.

وعليه فإن الدراسات السابقة قد تناولت استخدام هذه المعادلات على انفراد في تقدير الثبات ولم تقدم اي مقارنة بينها في ضوء بيانات حقيقية وعليه فان هذه الدراسة ستنفرد في تقديم مقارنة بين الطرائق الثلاثه في تقدير الثّبات في ضوء بيانات حقيقية باستخدام اختبارات تحصيلية ذات بعدين (نوعين من الفقرات).

7- منهج الدراسة وبياناتها:

تعتبر هذه الدراسة من الدراسات الوصفية إذ تتبع أسلوب المنهج الوصفي بشكل رئيس في وصف النتائج المتعلقة بمعاملات الثّبات والمقارنات بينها حيث استخدم (المتوسّط الحسابي والمدى والانحراف المعياري) في وصف تقديرات الثّبات كما استخدم الاحصاء التحليلي المتمثل باستخدام (اختبار كاي تربع وتحليل التباين الاحادي) للمقارنة بين معاملات الثّبات.

أما بيانات الدراسة فتكونت من نتائج الطلبة على الاختبارات النهائية (التي تطبق نهاية الفصل الدراسي) والتي يعدها أعضاء هيئة التدريس في الجامعات لتقويم أداء طلبتهم في المواد الدراسية، بواقع نتائج 84 اختباراً تحصيلياً تم اختيارها عشوائياً من الاختبارات النهائية للشعب الدراسية المطروحة في جامعة جدارا للفصل الدراسي الثاني 2016/2015 في مرحلة البكالوريوس استجاب على كل اختبار في المتوسّط حوالي 30 طالباً وطالبة وكان متوسّط عدد الفقرات حوالى 13 فقرة في كل اختبار وتكونت جميع الاختبارات من نوعين من الفقرات فقط هما: فقرات الاختيار من متعدد والفقرات المقالية (الإنشائية مصوغة الإجابة).

8- المعالجة الإحصائية :

تم تفريغ استجابات الطلبة على فقرات كل اختبار والبالغة حوالي 31980 استجابة (بالمتوسّط: 84 اختبار* 30 طالب * 13 فقرة)، وذلك بتفريغ استجابات كل شعبة على حدة في ملف spss، ثم بعد ذلك تقدير الثّبات لكل شعبة وفق معادلة كرونباخ ألفا (8) ومعادلة ألفا الطبقي (9) ومعادلة راجو

(7)، مع الأشارة إلى أن برمجية الـ spss لا توفر امكانية تقدير ألفا الطبقي وراجو وإنما توفر فقط امكانية تقدير ألفا، وعليه فقد تم استخدامها في حساب بعض الأجزاء الخاصة بمعادلتي راجو وألفا الطبقي. بعد ذلك تم حساب متوسّطات الثّبات المقدرة وفق كل طريقة وحساب قيمة اختبار كاي تربيع لمعرفة مدى التوافق بين الطرائق الثلاث في تقدير الثبات وحساب معامل التوافق بينها باستخدام معامل Contingency coefficient بالإضافة الى تحليل التباين الاحادي لمعرفة الفروق بين المتوسطات الثلاثة لمعاملات الثبات(ANOVA).

وفيما يتصل بتصنيف قيم معاملات الثّبات، وكونه لا يوجد قاعدة مطلقة للحكم عليها (Frisbie, 1988) وفي ضوء ما أشار له كل من (Frary, 2011; Feldt & Brennan, 1989) فقد صنّفت قيم معاملات الثّبات الناتجة في هذه الدراسة إلى ثلاثة مستويات هي: ثبات مرتفع، ثبات منخفض، وثبات متوسّط ، حيث اعتبر معامل الثّبات مرتفعا إذا زادت قيمة معامل الثّبات عن 0.70 ومتوسّطا إذا تراوح من 0.40 إلى 0.70 ومنخفضا إذا قلّ عن 0.40.

9– نتائج الدراسة:

للاجابة عن السؤال الاول" ما هي تقديرات الثّبات للاختبارات الجامعية التي تتكون من نوعين من الفقرات باستخدام كل من كرونباخ ألفا وألفا الطبقي وراجو؟" تم تقدير معاملات الثّبات وفق الطريقتين ويوضح الجدول1 قيم معاملات الثّبات المقدرة لكل اختبار وفق هاتين الطريقتين، كما يبين الإحصاءات الوصفية لمعاملات الثّبات المحسوبة مثل المتوسّطات الحسابية والوسيط والتباين والمدى لهذه القيم المحسوبة.

جدول 1

قيم معاملات الثّبات المقدرة لعينة الاختبارات وفق الطرائق الثلاث(ألفا، ألفا الطبقي، راجو).

رقم الاختبار	قيم معامل الثّبات Alpah	قيم معامل الثّبات Stratified alpha	قيم معامل الثّبات Raju	رقم الاختبار	قيم معامل الثّبات Alpah	قيم معامل الثّبات Stratified alpha	قيم معامل الثّبات Raju
1	0.63	0.72	0.33	43	0.88	0.94	0.77
2	0.65	0.69	0.69	44	0.82	0.87	0.11
3	0.88	0.89	1.29	45	0.63	0.74	0.26

قيم معامل الثّبات Raju	قيم معامل الثّبات Stratified alpha	قيم معامل الثّبات Alpah	رقم الاختبار	قيم معامل الثّبات Raju	قيم معامل الثّبات Stratified alpha	قيم معامل الثّبات Alpah	رقم الاختبار
0.53	0.8	0.80	46	0.50	0.65	0.56	4
0.42	0.85	0.78	47	0.56	0.75	0.63	5
0.17	0.71	0.65	48	0.65	0.76	0.65	6
0.15	0.9	0.76	49	0.43	0.60	0.53	7
0.56	0.83	0.75	50	0.40	0.56	0.48	8
0.77	0.9	0.79	51	0.77	0.88	0.81	9
0.69	0.89	0.65	52	0.40	0.76	0.66	10
0.50	0.79	0.75	53	0.50	0.86	0.76	11
0.62	0.85	0.77	54	0.77	0.75	0.65	12
0.75	0.86	0.78	55	0.78	0.87	0.79	13
0.51	0.90	0.82	56	0.87	0.9	0.88	14
0.40	0.79	0.69	57	0.63	0.87	0.84	15
0.66	0.78	0.70	58	0.66	0.83	0.74	16
0.59	0.69	0.63	59	0.66	0.72	0.67	17
0.78	0.82	0.75	60	0.20	0.83	0.74	18
0.10	0.73	0.64	61	0.78	0.79	0.67	19
0.52	0.87	0.80	62	0.44	0.79	0.68	20
0.94	0.88	0.86	63	0.46	0.65	0.62	21
0.62	0.88	0.83	64	0.33	0.91	0.67	22
0.35	0.83	0.73	65	0.71	0.80	0.74	23
0.42	0.79	0.72	66	0.10	0.63	0.47	24
-0.01	0.85	0.71	67	0.46	0.66	0.54	25
0.92	0.85	0.78	68	0.22	0.61	0.5	26
0.68	0.81	0.74	69	0.41	0.75	0.67	27
0.60	0.89	0.78	70	0.16	0.81	0.69	28
0.37	0.76	0.70	71	0.12	0.63	0.65	29
0.12	0.69	0.61	72	0.52	0.81	0.75	30
0.86	0.8	0.78	73	0.68	0.88	0.73	31

رقم الاختبار	قيم معامل الثّبات Alpah	قيم معامل الثّبات Stratified alpha	قيم معامل الثّبات Raju	رقم الاختبار	قيم معامل الثّبات Alpah	قيم معامل الثّبات Stratified alpha	قيم معامل الثّبات Raju
32	0.82	0.90	0.11	74	0.81	0.87	0.08
33	0.77	0.84	0.36	75	0.63	0.65	0.69
34	0.72	0.77	0.77	76	0.83	0.84	0.73
35	0.69	0.70	0.79	77	0.69	0.73	0.75
36	0.76	0.83	0.69	78	0.74	0.87	0.52
37	0.60	0.66	0.62	79	0.72	0.79	0.36
38	0.86	0.9	0.83	80	0.83	0.85	0.92
39	0.72	0.74	0.66	81	0.55	0.65	0.46
40	0.69	0.73	0.72	82	0.83	0.83	0.67
41	0.57	0.64	0.19	83	0.74	0.85	0.43
42	0.69	0.71	0.71	84	0.63	0.56	0.61

	الوسط الحسابي	أعلى قيمة	ادنى قيمة	المدى	الانحراف المعياري	معامل الالتواء
معامل الثّبات ألفا	0.71	0.88	0.47	0.41	0.09	0.47-
معامل الثّبات ألفا الطبقي	0.79	00.95	0.56	0.39	0.09	0.62-
معامل الثّبات راجو	0.53	0.98	0.05	0.93	0.24	0.11-

يلاحظ من الجدول 1 أن الوسط الحسابي لتقديرات الثّبات حسب معادلة ألفا قد بلغ (0.71)، و(0.79) حسب معادلة ألفا الطبقي و(0.53) حسب راجو. وعليه فإن مستوى تقديرات الثّبات كانت قيمه مرتفعة نسبياً وفق أسلوب ألفا الطبقي فيما تدنّت تقديرات معاملات الثّبات بشكل ملحوظ وفق راجو. وبذلك فإن أسلوب ألفا الطبقي يقدم تقديراً مرتفعاً للاختبارات المكوّنة من جزأين يليه أسلوب ألفا فيما يبخس أسلوب راجو الثّبات بشكل واضح نسبة لألفا وألفا الطبقي. كما ويلاحظ ضيق المدى في تقديرات الثّبات وفق أسلوبي ألفا وألفا الطبقي اذ بلغ (0.41 ، 0.39) على التوالي فيما ارتفع إلى (0.93) وفق أسلوب راجو .

أما فيما يتعلق بإجابة السوال الثاني "ما مدى التوافق بين كل من الطرائق الثلاثة (ألفا، راجو، وألفا الطبقي) في تقدير الثّبات لهذه الاختبارات التي تتكون من نوعين من الفقرات؟" فقد تم تصنيف معاملات الثّبات إلى مرتفع ومتوسّط ومنخفض حسب التصنيف المعتمد في هذه الدراسة المستند الى

دراسة كل من(Frary, 2011; Feldt & Brennan, 1989)، والجدول (2) يبين هذا التوزيع التكراري لمعاملات الثّبات وفق ذلك التصنيف.

جدول 2

توزيع قيم معاملات الثّبات وفق الطرائق الثلاث (ألفا وألفا الطبقي وراجو).

نوع معامل الثبات	مستوى معامل الثّبات وفق تصنيف (Frary, 2011)			المجموع
	مرتفع	متوسّط	منخفض	
ألفا	47 (55%)	37 (45%)	0 (0%)	84
ألفا الطبقي	67 (80%)	17 (20%)	0 (0%)	84
راجو	22 (26%)	41(49%)	21 (25%)	84

من الجدول أعلاه نلاحظ ارتفاعاً واضحاً جداً في اعداد قيم معاملات الثّبات التي صنّفت مرتفعة وفق معادلتي ألفا وألفا الطبقي نسبة إلى راجو الذي لم تتعدّ قيم الثّبات المرتفعة فيه إلى نصف ذلك الرقم فقد صنفت 80% من قيم معاملات الثبات بانها مرتفعة وفق معادلة ألفا الطبقي و 55% وفق معادلة ألفا فيما لم يصنّف سوى 26% منها مرتفعاً وفق معادلة راجو، كما نلاحظ عدم وجود أي معامل ثبات مصنّف على أنه منخفض وفق تقديرات ألفا وألفا الطبقي في حين كانت 25% من معاملات الثّبات وفق معادلة راجو منخفضة. ولمعرفة فيما إذا كان هناك فرقاً جوهرياً بين هذه التصنيفات أو التوزيعات لمعاملات الثبات وفق المعادلات الثلاثة، فقد تم حساب قيمة كاي تربيع χ^2 (مؤشر الاستقلالية) بين توزيعات معاملات الثّبات أعلاه ووجد أن قيمة كاي تربيع المحسوبة 11.01= χ^2 بدرجات حرية 4 وهي عبارة عن ((عدد الصفوف–1) x (عدد الاعمدة –1) وكانت قيمة الاحتمالية (p= 0.02) وهي دالة احصائياً على مستوى 0.05. وكون معامل الاستقلالية كاي تربيع يبين وجود علاقة أو عدمها فقط ولا يبيّن قوة هذه العلاقة فقد تم حساب معامل التوافق الكونتنجنسي Contingency وفق المعادلة (10):

$$\text{Contingency coefficient} = \text{square root}\left(\chi^2 / (\chi^2 + N)\right) \quad \ldots\ldots (10)$$

$$\text{Max (Contingency coefficient)} = \text{square root}(r-1/r) \quad \ldots\ldots (11)$$

وقد بلغت قيمته المحسوبة 0.36 وهي قيمة ضعيفة اذا ما قورنت باعلى قيمة يصلها وهي 1 وفق المعادلة (11) حيث في جدول 2 (عدد الصفوف = عدد الأعمدة =3).

بالإضافة لذلك تم استخدام تحليل التباين الاحادي ANOVA لمعرفة فيما إذا كانت هناك فروقا ذات دلالة احصائية بين متوسطات الثبات المحسوبة بالطرائق الثلاثة والجدول 3 يبين نتائج هذا التحليل.

جدول 3

تحليل التباين الأحادي لمعاملات الثّبات وفق الطرائق الثلاث (ألفا وألفا الطبقي وراجو).

مصدر التباين	درجات الحرية	مجموع المربعات	متوسط المربعات	قيمة الاختبار	الاحتمالية
بين المجموعات	2	2.940	1.470	55.597	.000
داخل المجموعات	249	6.583	.026		
الكلي	251	9.523			

تشير نتائج تحليل التباين أعلاه الى أن هناك فروقا ذات دلالة احصائية على مستوى 0.05 بين المتوسطات الثلاثة ($F(2,249)= 55.597 , P = 0$) كما وكشفت نتائج التحليل البعدي باستخدام اختبار Scheffé post hoc عن فروق دالة احصائياً بين كل زوج من المتوسطات ولصالح الوسط الأكبر في كل مرة.

11- مناقشة النتائج:

فيما يتعلق بنتائج السؤال الأول أظهرت نتائج الدراسة قيما متوسطة وبعضها مرتفعة لمعاملات الثبات المقدرة للاختبارات من اعداد المدرسين، وهذه القيم تقارب قيم الثبات التي أشار اليها فرسبي (Frisbie, 1988) في دراسته لتقدير ثبات الاختبارات من اعداد المدرسين والبالغة (0.5) كما تتفق مع ما توصلت له دراسة كينيوا واكيونيا (Kinyua, Okunya, 2014) في تقدير قيم الثبات للاختبارات من اعداد المدرسين Teacher Made Test والتي كانت متوسطة.

كما أظهرت النتائج اختلافاً في توزيع معاملات الثّبات بالطرائق الثلاثة، حيث بلغ متوسّط الثّبات لمعامل ألفا الطبقي (0.79) وهي قيمة مرتفعة حسب تصنيف برنان (Brennan,1989)، بينما بلغ متوسّط الثّبات لمعامل ألفا (0.71) وهي قيمة مرتفعة ايضاً، في حين بلغ متوسّط الثّبات للشعب الدراسية وفق تقديرات راجو فقط ما مقداره (0.53) وهي قيمة متوسّطة. كما وتشير النتائج بشكل واضح إلى أن أسلوب ألفا الطبقي كان يقدم تقديرات للثبات أعلى من غيرة من أساليب تقدير الثّبات باستمرار، وهذا يتفق مع (Cortina, 1993) بأن معامل الفا يبخس معامل الثبات نسبة إلى

غيره من المعاملات وخصوصا في حالة تعدد الابعاد للاختبار ويتفق ايضاً مع ما أشار له (Haertel, 2006) إلى أن معامل ألفا الطبقي يقدم تقديراً أفضل من غيره من المعاملات خصوصاً إذا كانت الارتباطات بين الفقرات ضمن النوع الواحد أكثر منها بين النوعين من الفقرات، وأن استخدام ألفا الطبقي أكثر دقة في تقدير الثّبات لأي اختبار يتضمن مجموعات جزئية من الفقرات تقيس محتوىً مختلفاً متعدد الأبعاد (Multidimensional) مقارنة بالعديد من معاملات الثّبات التي تم اشتقاقها ضمن شرط التوازي للأجزاء أو تشاكلها أو تكافؤها بالضرورة وهذا يتفق مع ما أشار له أسبورن (Osbourn, 2000) وكل من كاماتا وتورهان وداراندري (Kamata, Turhan ,Darandari, 2003) بأن ألفا الطبقي هو الأفضل، اذ يقدر ألفا الطبقي الثّبات للاختبار الكلي بتعامله مع علامة الاختبار كعلامة مركبة (Composite Score) من تركيب خطي لعلامتين فرعيتين أو أكثر. كما وتتفق قيم الثّبات المرتفعه المقدرة وفق كل من ألفا وألفا الطبقي مع دراسة كل من كوالز (Qualls, 199) ودراسة ميلر (miller's, 1998) واللتان أشارت نتائجهما إلى أن ارتفاع ثبات الاختبارات التحصيلية ذات النموذج المختلط الفقرات في الاختبارات الأمريكية قد تجاوز 0.80 . كما وتتفق أيضا نتائج هذه الدراسة مع دراسة (Qualls,1995) والتي أشارت دراسته إلى أن ثبات الاختبارات متعددة أنواع الفقرات في الاختبارات الأمريكية قد بلغت (0.802). كما وتتفق كذلك نتائج هذه الدراسة في تفوق قيم ألفا الطبقي مع دراسة كل من كاماتا وتورهان وداراندري (Kamata, Turhan, Darandari, 2003) والتي تفوق فيها معامل الثّبات ألفا الطبقي على أسلوب متعدد الأبعاد وأسلوب الثّبات الأقصى (Maximal Reliability, and Multidimensional omega).

12- التوصيات:

12- 1- استخدام ألفا الطبقي كبديل لكرونباخ ألفا في تقدير ثبات الاختبارات المركّبة من نوعين من الفقرات فأكثر وذلك لأن هذه الاختبارات تحقّق شروط التّشاكل ولا تحقق شروط التّوازي.

12- 2- إجراء دراسات مماثلة لتقدير الثّبات في حالة الاختبارات التي تتضمّن أكثر من نوعين من الفقرات، حيث اقتصرت هذه الدراسة على الاختبارات التي تتضمّن نوعين فقط من الفقرات.

12- 3- إجراء دراسة مماثلة للمقارنة بين هذه الطرائق مع الأخذ بعين الاعتبار أحادية البعد للاختبار.

12- 4- إجراء دراسة مماثلة لهذه الدّراسة على عينة أكبر من الاختبارات للتحقق من استقرار النتائج التي توصّلت لها هذه الدراسة.

المراجع العربية:

الشّايب، عبد الحافظ (2007). تقدير ثبات علامات عيّنة من المواد في جامعة آل البيت، *مجلة جامعة دمشق*، المجلد23، العدد الثاني، 255-271.

سوالمة، يوسف (1995). تقدير الثّبات للعلامات الجامعيّة في عيّنة من المساقات في جامعة اليرموك، *مجلّة مركز البحوث التربويّة بجامعة قطر*، السّنة الرابعة، العدد السّابع، 71-89.

المراجع الأجنبية

Brennan, R. L. (Ed.) (2006). *Educational measurement (4th Ed.).* Westport, CT. American Council on Education/ Praeger.

Cortina, J. M. (1993). *What is coefficient alpha*? An examination of theory and application

Crocer, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston,

Cronbach, L.J. (2004). *My current thoughts on coefficient alpha and successor procedures* (CSE Report 643). Los Angeles, Ca.: University of California.

Feldt, L. and Brennan, R. (1989). *Reliability. In educational Measurement* (3rd edition) Edited by R. Linn .New York: Macmillan Publishing Company.

Feldt, L. S. (1975). Estimation of the reliability of a test divided into two pans of unequal length. *Psychometrika, 40*, 557-561.

Frary, R. (2011). *Testing Memo 8*: Reliability of test scores. Retrieved from http://www.testscoring.vt.edu/memo08.html

Frisbie, D. (1988). Reliability of scores from teacher made tests. *Educational Measurement: Issues and Practice, 7*(1), 25-35.

Gleser, G., Cronbach, L., and Rajaratnam, N. (1965). Generalizability of scores influenced by multiple sources of variance. *Psychometrika, 30*(4):395-418. (Table 3, rearranged to show increasing patient severity and increasing item severity.

Guttman, L. (1945) A basis for analyzing test–retest reliability. *Psychometrika 10*:255–282

Haertel, E.H.(2005). *Reliability.in Educational Measurement*.4th ed,pp65-110.New York , American council on education and Macmillan.

Hogan, T.P., Benjamin, A. & Brezinski, K.L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement, 60*, 523-531.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*, 109- 133.

Kamata, A, Turhan, A and Darandari, E (2003) 'Scale Reliability of Multidimensional Composite Scores'. *Paper presented at the annual meeting of American Educational Research Association*, Chicago, April 2003.

Kamata, A., Turhan, A., & Darandari, E. (2003). *Estimating reliability for multidimensional composite score scale scores*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Keng, Miller, O'Malley, Turhan. (2005). *A Generalization of Stratified α that Allows for Correlated Measurement Errors between Subtests*. Pearson, 400 Center Ridge Drive, Austin.

Retrieved from:
http://images.pearsonassessments.com/images/tmrs/tmrs_rg/StratifiedAlphathatAllowsforCorrelatedMeasurementErrorsbetweenSubtests.pdf?WT.mc_id=TMRS_A_Generalization_of_Stratified.

Kinyua, K; Luke Okunya,O. (2014). Validity and reliability of teacher-made tests: Case study of year 11 physics in Nyahururu District of Kenya. *African Educational Research Journal. 2* (2), 61-71.

Lord, F. M. & Novick, R. (1968). Statistical theories of mental test scores. Reading MA: *Measurement, 53*, 33-49.

Murphy, K. R. & DeShon, R. P. (2000). Inter-rater correlations do not estimate the reliability of nonhomogeneous items. *Applied Psychological Measurement*, *22*, 375-385.

Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika, 32*, 1–13.

Osbourn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods, 5*, 343-355.

Qualls, L. A. (1995). Estimating the Reliability of a Test Containing Multiple Item Formats. *Applied Measurement in Education. 8*(2), 111-120.

Raju, N. S. (1977). A generalization of coefficient alpha. *Psychometrika, 42*, 549-565.

Raykov, T. (1998). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement, 22*, 375–385.

Rulon, PJ. (1937). A simplified procedure for determining the reliability of a test by split-halves. *Harvard Educational Review 9*:99-103.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. Psychological Assessment, 4, 350-353. test reliability under violation of two assumptions. Educational and Psychological.

Warrens, J. M. (2013). *Cronbach's alpha as the mean of all split-half reliabilities*. Retrieved from http://www.matthijswarrens.nl/talks/13_07_25.pdf

Zimmerman, D. W., Zumbo, B. D., & Lalonde, C. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions. *Educational and Psychological Measurement, 53*, 33-49

<< وصل هذا البحث إلى المجلة بتاريخ 2016/2/9 وصدرت الموافقة على نشره بتاريخ 2017/4/6>>